



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Data warehouse til elbilers opladning og elpriser

Andersen, Ove; Krogh, Benjamin Bjerre; Torp, Kristian

Publication date:
2014

Document Version
Tidlig version også kaldet pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Andersen, O., Krogh, B. B., & Torp, K. (2014). *Data warehouse til elbilers opladning og elpriser*. Institut for Datalogi, Aalborg Universitet. <http://www.ens.dk/elbilers-energiforbrug>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Data warehouse til elbilers opladning og elpriser

Daisy, Institut for Datalogi

Aalborg Universitet

Teknisk rapport

December 2014

Dokument titel:	Data warehouse til elbilers opladning og elpriser
Bevillingsmyndighed	Energistyrelsen
Version:	1.0
Publiceringsdato:	22. december 2014
Type af dokument:	Rapport
Synlighed:	Offentlig
Forfattere:	Ove Andersen, Benjamin B. Krogh, Kristian Torp
Ansvarlig for rapport	Kristian Torp, torp@cs.aau.dk
Organisation	Daisy, Institut for Datalogi, Aalborg Universitet

Resumé

Denne rapport præsenterer, hvordan GPS og CAN bus målinger fra opladning af elbilerne er rensset for typiske fejl og gemt i et data warehouse. GPS og CAN bus målingerne er i data warehouse integreret med priserne fra det Nordeuropæiske el spotmarked Nord Pool Spot. Denne integration muliggør, at man kan analysere på priser for at oplade en elbil samt vurdere på, hvordan disse opladninger kan indpasses i el infrastrukturen.

GPS og CAN bus målingerne er opsamlet fra 176 elbiler (Citroen C-Zero, Mitsubishi iMiev og Peugeot Ion) i kalenderårene 2012 og 2013 og målingerne stammer fra Test-en-elbil projektet. Data er rensset for fejl så som uventede hop i kilometertælleren, uventede hop i State of Charge (SoC) værdien og svingende SoC værdier. For hver fejltypen er fejlen beskrevet med et konkret eksempel fra rå data og antal af fejl er listet. Det beskrives yderligere, hvordan hver enkelt fejl er håndteret inden data læses ind i data warehouse.

Elbiler kan oplades både hjemme og ude. Elbilerne anvendt i Test-en-elbil projektet skifter bruger hver 3.-6. måned. GPS data anvendes til at bestemme, hvor en bil hører hjemme. Herefter klassificeres opladninger som værende hjemme eller ude.

GPS og CAN bus målinger om opladningen af elbiler er sammen med priserne fra el spotmarkedet indlæst i et data warehouse, som er fuldt ud implementeret. Den logiske data model for dette data warehouse præsenteres i detaljer. Håndteringen af GPS og CAN bus målingerne er generisk og kan udvides til nye data kilder inklusive data fra andre lande.

Summary

This report presents how GPS and CAN bus data from charging electric vehicles (EVs) is cleansed for typical errors and stored in a data warehouse. To enable analysis on the price for charging an EV the data warehouse also contains price information from the Nord Pool Spot electricity spot market.

The GPS and CAN bus data is collected from 176 EVs (Citroen C-Zero, Mitsubishi iMiev, and Peugeot Ion) during the years 2012 and 2013 in the Danish EV project named Test-en-elbil (Try-an-EV). Typical error such as unexpected jumps in the odometer value, unexpected jumps in State-of-Charge (SoC), and oscillating SoC values are handled. For each type of error, an example is provided along with the number of errors found. In addition, it is shortly explained how the errors are dealt with.

EVs can be charged both at home and in the city. The vehicles used in the project Test-en-elbil change user every 3 to 6 months. The GPS data is used to determine when a vehicle is at home. This information is used to classify each charging of the EV as at home or away from home.

The GPS and CAN bus data related to charging the EVs along with the pricing information is integrated into a data warehouse that is fully implemented. The logical data model for the data warehouse is described in details. The handling of GPS and CAN bus data is generic and can be adopted to handle EV charging data from other sources and in other countries.

Indholdsfortegnelse

Resumé	3
Summary	3
Liste af figurer	6
Liste af tabeller	6
Termer	7
1. Introduktion	8
2. Datagrundlag	9
2.1 GPS og CAN bus data	9
2.2 Elpriser	9
3. Metode	12
3.1 Rensning af opladningsdata fra GPS og CAN bus data	12
3.1.1 Kilometertæller	12
3.1.2 State of Charge (SoC) /Opladningsprocent	15
3.2 Rensning af spotpris data	17
3.3 Beregninger på opladningsdata	18
3.3.1 Brugere og hjemmeladere	18
3.3.2 Ladestationer	21
3.3.3 Opladningsintervaller og priser	21
4. Logisk data-warehouse model	24
4.1 Model	24
4.2 Dimensioner	24
4.2.1 Dimdate	25
4.2.2 Dimtime	25
4.2.3 Dimvehicle	25
4.2.4 Dimuser	25
4.2.5 Dimdatasource	25
4.2.6 Dimspotmarket	25
4.2.7 Dimspotprice	25
4.2.8 Dimchargingstation	25
4.2.9 Dimchargingslot	26
4.3 Fact tabeller	26
4.3.1 Factchargedata	26
4.3.2 Factchargestats	26

4.4	Bridge tabel	26
4.4.1	Bridgechargespotprice.....	26
4.5	Implementation.....	27
5	Konklusion	28
	Litteraturliste.....	29
	Bilag 1	30

Liste af figurer

Figur 1 Gennemsnitlig daglig elpris for 2012 og 2013.	10
Figur 2 Gennemsnitlig pris i DKK/kWh på timeinterval for DK1 og DK2 over 2012 og 2013.	10
Figur 3 Gennemsnitlig pris pr. ugedag i DKK/kWh på timeinterval for DK1 og DK2 over 2012 og 2013	11
Figur 4 Antal observationer i forhold til antal kilometer kilometertælleren hopper (0-100 km).	13
Figur 5 Antal SoC spring i forhold til hvor mange SoC der springers over.....	16
Figur 6 Ladningsstationer og deres aktive intervaller fra Tabel 9.	19
Figur 7 Faktiske opladning udført for hver ladestation vist i Figur 6.....	19
Figur 8 Ladestationer sorteret efter Benyttelsesgraden, som vist i Tabel 10. Ladestationerne er farvet efter brugere.	21
Figur 9 Visualisering af pris-, plugin- og opladningsperioder.	21
Figur 10 Visualisering af tilpassede opladningsperioder, individuelle opladningsprocenter og ventetid før opladningsstart.	22
Figur 11 Logisk model for opladningsdata warehouse.....	24

Liste af tabeller

Tabel 1 Gennemsnitlige elpriser og karakteristika for DK1 og DK2.....	9
Tabel 2 Kilometertæller springer.....	12
Tabel 3 Kilometertæller springer til over 4 milliarder.	13
Tabel 4 Eksempel på enkeltstående fejlværdi.....	14
Tabel 5 Eksempel på tidsspring.	14
Tabel 6 Sortering af Tabel 5 efter tidspunkt.....	15
Tabel 7 Eksempel hvor opladningsprocent springer procenter over.	15
Tabel 8 Eksempel på at opladningsprocent fluktuerer mellem 99 og 100 %.	17
Tabel 9 Ladestationer benyttet for køretøj med id VF31NZKZZBU801081, sorteret efter første dato.	18
Tabel 10 Ladestationer benyttet af et køretøj (ID=VF31NZKZZBU801081) i 2012, sorteret efter benyttelsesgrad.	20
Tabel 11 Størrelse af tabeller i antal rækker og bytes.....	27
Tabel 12 Enterprise Bus Matrix af data warehouse modellen fra Figur 11.....	27

Termer

Term	Beskrivelse
Beskidt data	Data der indeholder overraskende værdier så som en bil, der kører med en meget høje hastigheder, f.eks. 300 km/t, eller meget lave temperaturer f.eks. – 50 grader celsius. Beskidt data forefindes i praksis altid i større datasæt og kan skyldes f.eks. midlertidig apparatfejl eller fejl under kommunikation.
Benyttelsesgrad	En 1:1 vægtning mellem hvor mange dage der er mellem første og sidste opladning ved en opladestation samt antallet af opladninger ved en opladestation.
CAN bus data	Controller Area Network (CAN) bus måling, der indeholder værdier for energiforbrug, batterikapacitet tilbage og andre relevante værdier.
Data warehouse	Data model meget velegnet til at gemme og forespørge på historisk data.
GPS data	Måling der indeholder breddegrad, længdegrad, kompasretning og andre relevant værdier.
Hurtig ladestation	En oplader der er defineret i data som "Quick Charge" og som er en DC oplader med effekt på op til 50 kWh [1].
Måling	En række fra en CSV fil med kolonner indeholdende GPS og CAN bus data værdier.
SoC	State of Charge (SoC) er batteriniveau i procenter f.eks. 25 SoC er lige 25 procent af batterikapaciteten er tilgængelig.
Standard ladestation	Når en ladestation ikke er en hurtig ladestation er den defineret som "Onboard Charger", altså en AC oplader med en effekt på op til 3,7 kWh. Bemærk at standard ladestationer både kan være i hjemmet og i byen. [1]

1. Introduktion

Denne rapport beskriver, hvordan opladningsdata udtrukket fra GPS og CAN bus målinger fra elbiler først er rensset for typiske fejl og dernæst gemt i et data warehouse. Dette data warehouse integrerer opladningsdata med elpriserne fra Nord Pool Spot spotmarkedet for el.

GPS og CAN bus data målingerne anvendt i rapporten er fra Test-en-elbil projektet [2]. Der er i alt målinger fra 176 elbiler. Disse biler er en blanding af de tre biler Citroen C-Zero, Mitsubishi iMiev og Peugeot Ion og kaldes populært "trillingerne" fordi de i praksis er ens [2]. Data er opsamlet i kalenderårene 2012 og 2013. De historiske elpriser er hentet fra Nord Pool Spots hjemmeside [3].

Det er meget interessant at integrere opladningsdata for elbiler med spotprisen på el fordi disse elpriser generelt er lavere om natten og i disse perioder, med billig strøm, holder elbiler typiske stille og er tilsluttet elnettet for at blive opladet til næste dags kørsel. Der kan derfor være store økonomiske og produktionsmæssige fordele ved at oplade elbiler om natten eller på andre tider med billig strøm.

En meget typisk måde at integrere historisk information, så som opladningsdata og elpris data, er via et såkaldt data warehouse [4]. Hovedfordelen ved et data warehouse er, at det er let at lave historiske forespørgsler så som "hvor ofte oplader en elbil?" eller "hvad er gennemsnitsprisen for at oplade en elbil i byen onsdag eftermiddage i sommerhalvåret?". En udfordring ved at anvende et data warehouse er, at det er nødvendigt at rense data for typiske fejl. Denne rensning kan være både tidskrævende og teknisk udfordrende. Det vurderes, at fordelene ved et data warehouse meget klart overstiger ulemperne. Denne rapport præsenterer derfor i detaljer, hvordan GPS og CAN bus målinger er rensset for typiske fejl, integreret med pris data og læst ind i et data warehouse.

Data warehouset er beskrevet i detaljer, hvor første den logiske model er præsenteret. Denne model giver et overblik. Præsentationen af modellen er teknisk orienteret, men forsøgt holdt simpel ved at præsentere modellen i mindre komponenter af gangen. Herudover er der i den tekniske beskrivelse referencer til litterature for de interesserede læsere. Som det næste er de enkelte komponenter i data warehouset beskrevet i detaljer.

Rapporten er opbygget som følgende. I afsnit 2 beskrives den rå data som rapporten bygger på. Dette er hhv. GPS og CAN bus målingerne samt de historiske elpriser. Afsnit 3 beskriver først, hvordan rå data er rensset for typiske fejl. Herefter beskrives, hvordan brugere og hjem er identificeret. Afsnit 4 er meget tekniske orienteret og præsenterer først den logiske data warehouse model, hvor opladningsdata og elpriserne er integreret. Der er herefter en detaljeret gennemgang af de enkelte komponenter i data modellen. Afsnit 5 konkluderer rapporten.

2 Datagrundlag

Data der anvendes i denne rapport er fra opladninger af elbiler foretaget i projektet Test-en-elbil [2]. Data er opsamlet fra 176 elbiler (Citroen C-Zero, Mitsubishi iMiev og Peugeot Ion) i kalenderårene 2012 og 2013. I disse data er der både oplysninger om elbilers energiforbrug ved kørsel og hvordan elbilerne oplades.

Spotpris på el på kommercielle markeder såsom Nord Pool Spot [3] kan varierer betydeligt. Elprisen har naturligvis en indflydelse på, hvad prisen er for at oplade en elbil, derfor er prisen fra Nord Pool Spot integreret med opladningsdata fra elbiler. Der vil i denne rapport udelukkende blive set på elpriser på spotmarkedet, disse priser er uden afgifter og leveringsomkostninger og er betydeligt lavere end de priser en privat forbruger skal betale.

2.1 GPS og CAN bus målinger

Rapporten er baseret på godt 216 millioner GPS og CAN bus målinger opsamlet fra de 176 elbiler i 2012 og 2013. I disse målinger er der både oplysninger om, hvad energiforbruget er ved kørsel og hvordan elbilerne oplader. I denne rapport fokuseres der udelukkende på, hvordan elbiler oplades. Oplysninger om elbilers forbrug ved kørsel, baseret på de samme målinger, kan findes i eksisterende arbejde [5] [6] [7] [8] [9].

Det er tidligere i detaljer beskrevet, hvordan GPS og CAN bus målinger modtages, læses, renses og kvalitetsbestemmes [5] [6]. Når disse rensede GPS og CAN bus målinger benyttes til analyser af opladninger læses disse ud for hvert køretøj individuelt og behandles i sorteret rækkefølge efter tidspunktet.

Selvom GPS og CAN bus målingerne er blevet rensset, er de tidligere renseregler udelukkende udarbejdet for at kunne lave analyser på elbilers forbrug ved kørsel. Det er derfor nødvendigt lave yderligere renseregler for at kunne anvende GPS og CAN bus målingerne til analyser på opladningsdata. Som et eksempel er der ikke i eksisterende renseregler taget hensyn til kvaliteten af data, når et køretøj er parkeret, da disse data ikke har været benyttet i det tidligere arbejde [5] [7]. Herudover inddrages der ekstra data fra GPS og CAN bus målingerne for at kunne lave analyserne på opladninger. Den ekstra data er f.eks. tidspunkter, koordinater, kilometertæller, gear (som også beskriver tilslutning til elnettet) [1] samt lufttemperaturen når en elbil oplades [9].

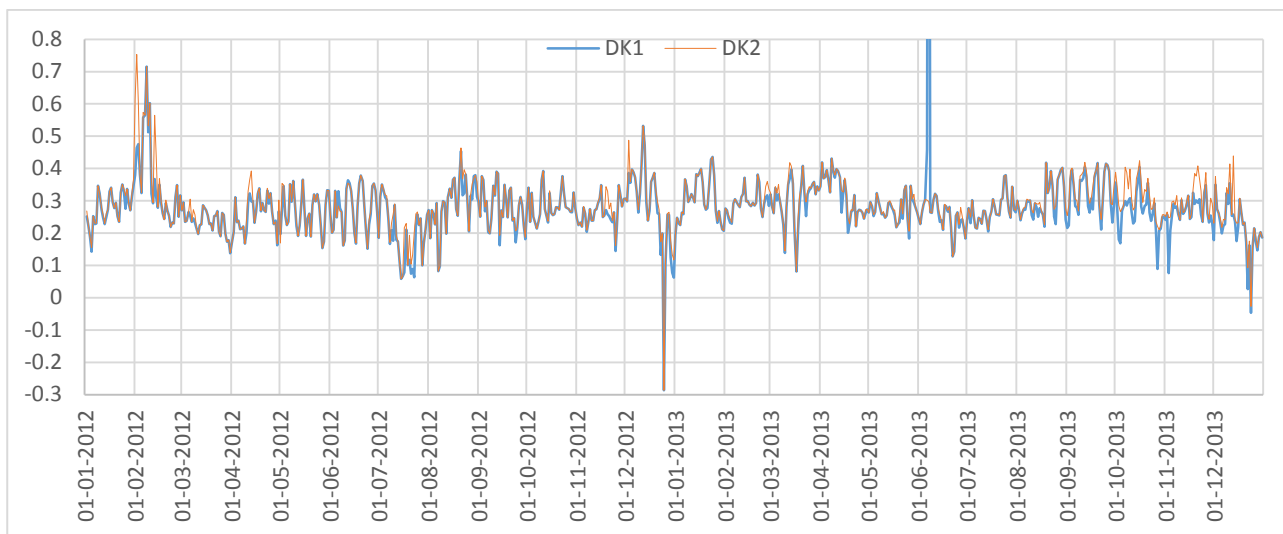
2.2 Elpriser

Historiske elpriser er tilgængelige fra Nord Pool Spots prisdatabase [3], som indeholder historiske elpriser med en granularitet på en time (24 værdier på et døgn) fra 2012 og frem. Danmark er inddelt i to områder, nemlig DK1, som er Jylland/Fyn, og DK2 som er Sjælland.

Spotmarked 2012 + 2013	Gennemsnit elpris (DKK/kWh)	Standardafvigelse (DKK/kWh)	Minimum (DKK/kWh)	Maksimum (DKK/kWh)	Negativ pris (timer)
DK1 (Jylland/Fyn)	0,267	0,230	-1,492	14,911	294
DK2 (Sjælland)	0,274	0,108	-1,492	1,887	225

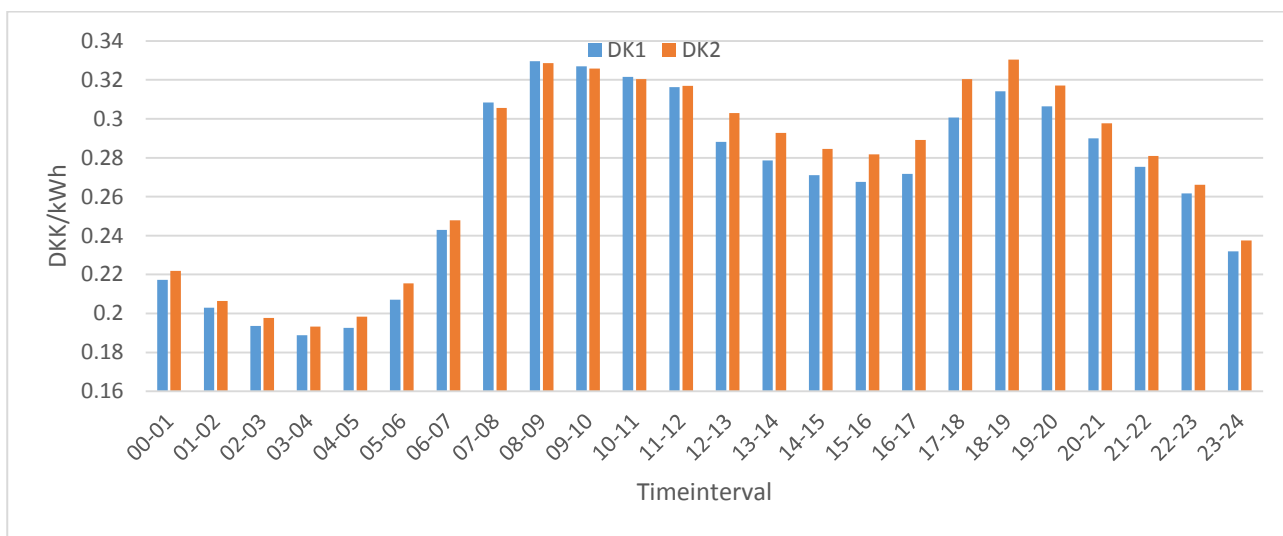
Tabel 1 Gennemsnitlige elpriser og karakteristika for DK1 og DK2.

Tabel 1 viser generelle pris oplysninger for de to geografiske elforsyningsområder, der dækker Danmark, for hele 2012 og 2013. Ud fra Tabel 1 kan det ses, at gennemsnitsprisen for DK1 er 0,07 DKK/kWh billigere end for DK2, mens at prisen i DK2 er meget mere stabil, med en standardafvigelse på blot 0,108 DKK/kWh mod 0,23 for DK1. Den laveste elpris for DK1 og DK2 er den samme, nemlig -1,492 DKK/kWh. Det vil sige, at i den givne time, har købere af strøm tjent 1,492 DKK/kWh. Negative priser, hvor forbrugeren har kunnet tjene på at bruge strøm, er der registreret i 294 timer i DK1 og 225 timer i DK2. Som forklaring på negative priser på strøm se f.eks. [10].



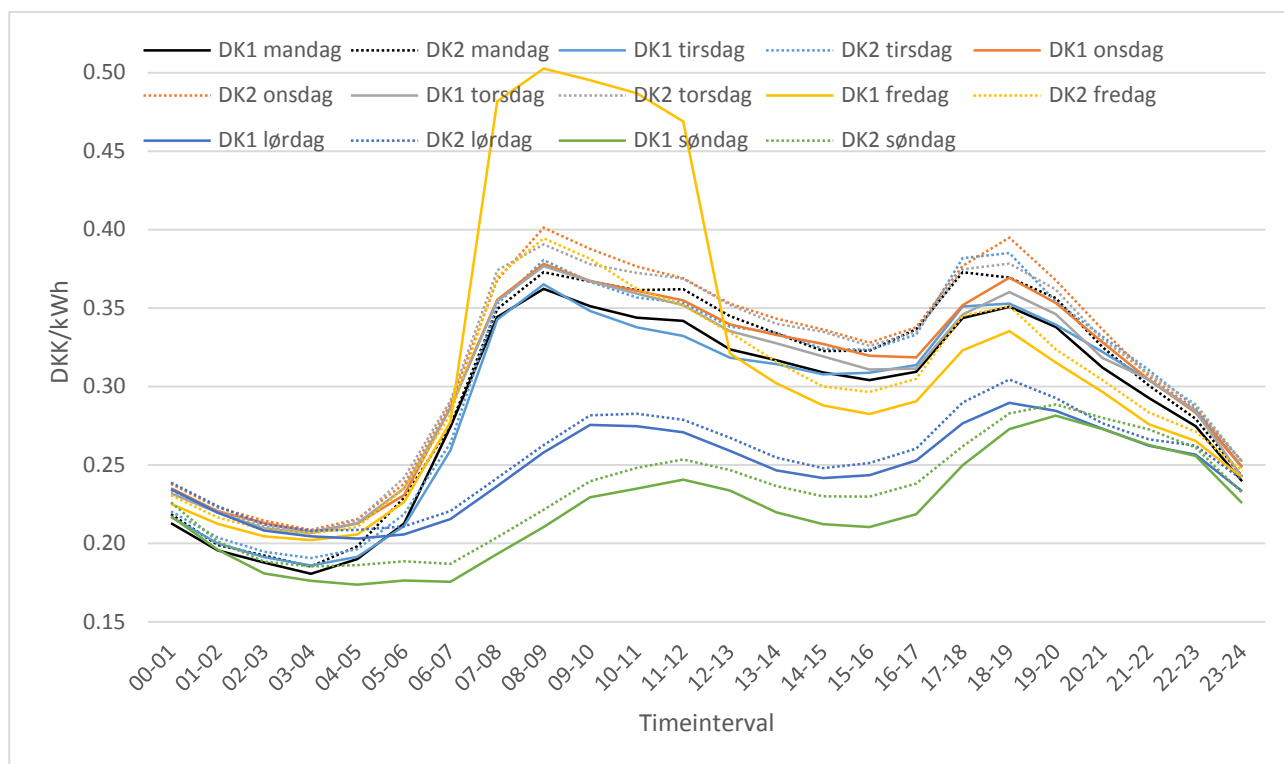
Figur 1 Gennemsnitlig daglig elpris for 2012 og 2013.

Figur 1 viser den gennemsnitlige daglige elpris for 2012 samt 2013 for hhv. DK1 (blå) og DK2 (orange). Prisen ligger oftest og svinger mellem 0,2 og 0,4 DKK/kWh, altså svinger prisen dagligt med en faktor 2. Der er også større udsving f.eks. er prisen over 0,5 DKK/kWh både i februar 2012, den 31. december 2012 samt den 7. juni 2013, hvor den gennemsnitlige dagspris for DK1 var på 3,25 DKK/kWh. Ligeledes er der dage med meget lave elpriser, eksempelvis den 25. december 2012, hvor den gennemsnitlige elpris var under minus 0,28 DKK/kWh for både DK1 og DK2, altså negativ strømpris.



Figur 2 Gennemsnitlig pris i DKK/kWh på timeinterval for DK1 og DK2 over 2012 og 2013.

Figur 2 viser den gennemsnitlige pris over døgnet for hhv. DK1 (blå) og DK2 (orange). Her ses det, at elprisen generelt er betydeligt lavere om natten. Mellem 00 og 06 er gennemsnitsprisen under 0,22 DKK/kWh, mens prisen mellem 07 og 12 og igen mellem 17 og 19 altid er over 0,3 DKK/kWh.



Figur 3 Gennemsnitlig pris pr. ugedag i DKK/kWh på timeinterval for DK1 og DK2 over 2012 og 2013

For at differentiere mellem ugedage viser Figur 3 de enkelte ugedages elpriser for DK1 og DK2. Her viser de hele linjer priserne for DK1, mens de stiplede linjer viser priserne for DK2. Det ses tydeligt, at priserne i weekenderne, blå og grøn, ligger betydeligt lavere end priserne på hverdagene. Hverdagspriserne følger hinanden jævnt, bortset fra DK1 om fredagen, hvor priserne springer til over 0,45 DKK/kWh mellem 7 og 12, hvor resten af hverdagene for både DK1 og DK2 i denne periode har en pris på 0,35 til 0,4 DKK/kWh.

3 Metode

I dette afsnit vil det blive beskrevet, hvordan data er håndteret og hvilke renseregler, der er implementeret for at tage højde for uventede værdier i opladningen af elbiler. Ligeledes vil forskellige metoder blive beskrevet til, hvordan brugere genkendes, hvordan ladestationer findes og hvordan den faktiske og den optimale priser for hver opladning beregnes.

3.1 Rensning af opladningsdata fra GPS og CAN bus målinger

De følgende afsnit beskriver, hvilke renseregler, der er brugt for at tage højde for uventede værdier i GPS og CAN bus målingerne. Bemærk at alle eksempler på uventede værdier fundet i data, der vises i det efterfølgende, er taget direkte fra de filer vi har modtaget. Dette afsnit beskriver renseregler ud over de renseregler, der anvendes for at bruge GPS og CAN bus målinger til at analysere på energiforbruget ved kørsel. For disse eksisterende renseregler se f.eks. [6].

3.1.1 Kilometertæller

Til analyser af opladninger er det relevant at vide hvor mange kilometer, der er kørt siden sidste opladning samt det gennemsnitlige forbrug siden sidste opladning. Ved en gennemgang af data har det vist sig, at der er egenskaber ved kilometertælleren, man skal tage højde for, som gør antal kilometer kørt besværligt at beregne.

3.1.1.1 Spring

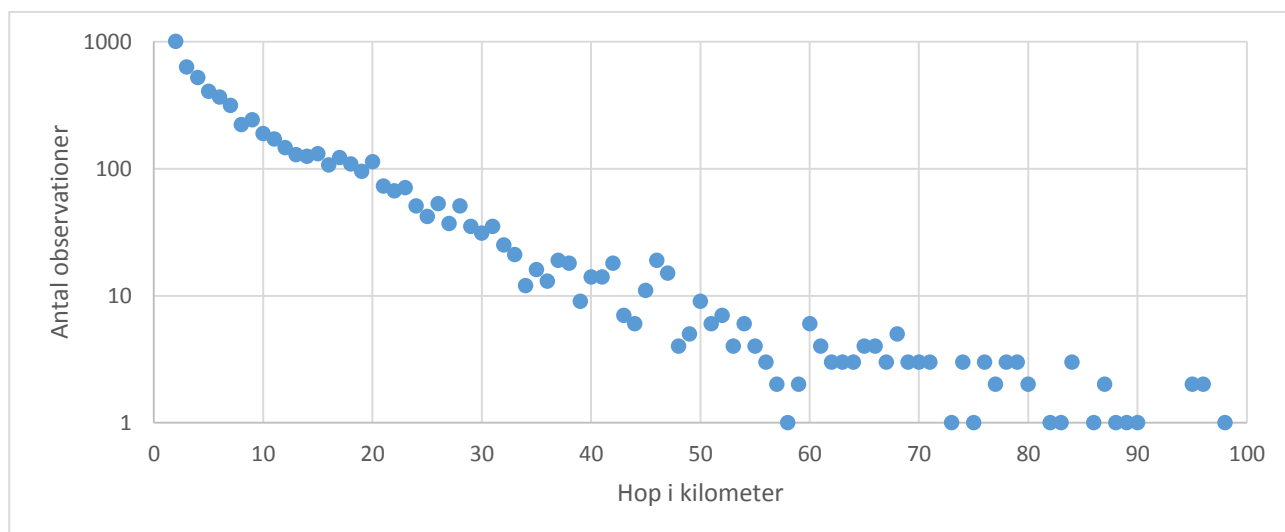
Det sker relativt ofte, at kilometertælleren springer et antal kilometer, så der bliver "huller" i sekvensen af kilometermålinger.

Køretøjs id	UTC tidspunkt	Kilometertæller
JMBLDHA3WBU001618	2013-01-10 11:00:15,945	24.103
JMBLDHA3WBU001618	2013-01-10 15:11:53,305	24.103
JMBLDHA3WBU001618	2013-01-10 16:28:39,530	24.173
JMBLDHA3WBU001618	2013-01-10 16:28:41,532	24.173

Tabel 2 Kilometertæller springer.

Et eksempel for køretøjet med id *JMBLDHA3WBU001618* er vist i Tabel 2. Her er vist et udsnit af en datafil med fire rækker af data, der kommer den viste rækkefølge. Her kan det ses, at fra 15:11:53,305 til 16:28:39,530 er der ikke registreret målinger, men alligevel er kilometertælleren hoppet fra 24.103 til 24.173, altså et spring på 70 km.

Dette spring kan findes ved at sammenligne GPS koordinaterne mellem 15:11:53,305 og 16:28:39,530 (koordinater ikke vist her a.h.t. anonymitet), at køretøjet har flyttet sig 58 kilometer i euklidisk afstand. Altså har logningsudstyret været frakoblet under visse ture, muligvis hvis et køretøj fragtes mellem to brugere.



Figur 4 Antal observationer i forhold til antal kilometer kilometertælleren hopper (0-100 km).

Fænomenet med at kilometertælleren springer er registreret i alt 6.268 gange, hvor 6.059 af disse observationer er vist i Figur 4. Bemærk venligst, at y-aksen med antal observationer er logaritmisk. Her kan det ses, at langt de fleste kilometerhop er på relativt få kilometer, men der er også registreret 96 spring på mellem 1.000 og 25.591 kilometer.

Udfordringen med springende kilometertællere er håndteret ved, at når der registreret, at der er kørt mere end 200 kilometer på en opladning, da vil kilometertallet blive sat til nul (0), således at denne opladning ikke kan bruges til f.eks. at beregne kWh/km, men opladningen kan stadig anvendes til f.eks. at se på, at opladningen startede ved 20 SoC og endte ved 100 SoC.

3.1.1.2 Ekstreme værdier

Når en kilometertæller i en bil når til tallet 32.767 sker det, at det næste registrerede kilometertal ikke er 32.768 men der imod 4.294.934.528.

Køretøjs id	UTC tidspunkt	Kilometertæller
JMBLDHA3WBU001738	2013-01-12 11:29:34,379	32.767
JMBLDHA3WBU001738	2013-01-12 11:29:35,380	32.767
JMBLDHA3WBU001738	2013-01-12 11:29:36,381	4.294.934.528
JMBLDHA3WBU001738	2013-01-12 11:29:37,382	4.294.934.528
JMBLDHA3WBU001738
JMBLDHA3WBU001738	2013-01-12 11:30:37,471	4.294.934.529

Tabel 3 Kilometertæller springer til over 4 milliarder.

Tabel 3 viser et eksempel, hvor kilometertælleren i løbet af 1 sekund springer fra 32.767 til 4.294.934.528. Dette virker umiddelbart som et ganske stort spring, men ved nærmere undersøgelse viser der sig et sammenhæng mellem disse to tal. Dette sammenhæng er analyseret i Bilag 1. Bemærk at når først en kilometertæller er hoppet til 4.294.934.528 tælles der videre fra denne værdi, som ses i sidste række i Tabel 3, hvor tælleren er hoppet til 4.294.934.529.

Der er fundet i alt 17 køretøjer i datasættet, hvor kilometertælleren efter 32.767 springer til en værdi over 4 milliarder og fortsætter derfra og i alt er der 2.716.024 målinger med et kilometertal over 4 milliarder.

Problemet med den springende kilometertæller er håndteret ved, at den høje kilometertællerværdi omregnes til en estimeret kilometertællerværdi som beskrevet i Bilag 1.

3.1.1.3 Enkeltstående fejlværdier

Kilometertælleren rapporterer for hver måling det kørte antal kilometer et køretøj har tilbagelagt totalt, men der optræder enkeltstående fejlværdier i denne data.

Køretøjs id	UTC tidspunkt	Kilometertæller
VF71NZKZZBU900326	2013-02-03 10:31:44,002	23.450
VF71NZKZZBU900326	2013-02-03 10:32:44,114	23.450
VF71NZKZZBU900326	2013-02-03 10:32:45,116	23.501
VF71NZKZZBU900326	2013-02-03 10:33:26,192	23.450
VF71NZKZZBU900326	2013-02-03 10:33:26,192	23.450

Tabel 4 Eksempel på enkeltstående fejlværdi.

Tabel 4 viser et eksempel for køretøjet med id VF71NZKZZBU900326, hvor køretøjet har rapporteret fem kilometertal, men hvor de to første og to sidste målinger har en kilometertæller værdi på 23.450 har den midterste en værdi på 23.501 kilometer. Det vides ikke, hvorfor sådan værdier optræder og der er ikke fundet nogen sammenhæng i afvigelserne af målingerne. Her springer kilometertælleren først +51 kilometer og dernæst falder den igen med -51 kilometer.

Sådanne enkelte spring i data er registreret i alt 3.576 gange, hvor der er registreret 1.635 enkeltstående spring i negativ retning, hvor en enkeltstående måling er lavere end målingen før og efter og er registreret 1,941 gange, hvor en enkeltstående måling er højere end målingen før og efter.

Disse fejlværdier håndteres ved, at rækken ikke anvendes i data warehouset. Dette har begrænset indvirkning, da det kun er individuelle sekundlogs, der fjernes. Dette svarer til, at der mangler en enkelt måling.

3.1.1.4 Tidsspring

Tiden er essentiel, fordi det er tiden, der bestemmer i hvilken rækkefølge GPS målingerne er optaget (den rækkefølge GPS målinger sorteres i). Derfor kan det have u hensigtsmæssige sideeffekter, hvis tidspunkter ikke er korrekt eller ikke er kronologiske.

Køretøjs id	Række nr.	UTC tidspunkt	Kilometertæller
VF31NZKZZBU801081	57	2013-01-11 07:42:29,885	21.038
VF31NZKZZBU801081	58	2013-01-11 07:42:30,886	21.038
VF31NZKZZBU801081	59	2013-01-11 07:41:31,916	21.038
VF31NZKZZBU801081
VF31NZKZZBU801081	116	2013-01-11 07:42:29,002	21.039
VF31NZKZZBU801081	117	2013-01-11 07:42:30,005	21.039

Tabel 5 Eksempel på tidsspring.

Tabel 5 viser tidspunkter og kilometertæller for køretøj med id VF31NZKZZBU801081. Her kan det ses, at mellem række 58 og 59 springer tiden 58,97 sekunder tilbage i tiden. Dette bevirker, at når data senere sorteres efter tidspunkter, da vil rækkefølgen af data målinger blive ændret således, at række 59 vil komme først.

Række nr.	Tidspunkt	Kilometertæller
59	2013-01-11 07:41:31,916	21.038
116	2013-01-11 07:42:29,002	21.039
57	2013-01-11 07:42:29,885	21.038
117	2013-01-11 07:42:30,005	21.039
58	2013-01-11 07:42:30,886	21.038

Tabel 6 Sortering af Tabel 5 efter tidspunkt.

Tabel 6 viser Tabel 5 sorteret efter tidspunkt. Det kan her ses, at rækkefølgen nu er væsentlig anderledes. Bemærk at kilometertælleren nu også skifter op og ned mellem 21.038 og 21.039. Dette kan give problemer flere steder, da GPS positionerne også vil hoppe frem og tilbage.

Det er fundet i alt 2.215 gange, hvor tiden springer tilbage og det berører. I alt 65.934 rækker med målinger der overlapper i tid, dvs. tiden springer tilbage og den samme periode har GPS målinger to gange, grundet tidsspring.

Håndtering af disse spring i tiden sker ved, at når data indlæses forventes det, at rækkerne er sorteret i tidsmæssig orden. Læses der tidspunkter, hvor tiden er sat tilbage, da ignoreres disse rækker som værende fejl indtil, at tiden er senere end da tidsspringet startede.

3.1.2 State of Charge (SoC) /Opladningsprocent

Opladningsprocenten (kaldet State of Charge eller forkortet SoC) er den grundlæggende enhed for måling af energiniveauet på et batteri samt hastighed med hvilken en opladning foregår. Derfor er det vigtigt at kende SoC så nøjagtigt som muligt.

3.1.2.1 Spring i værdier

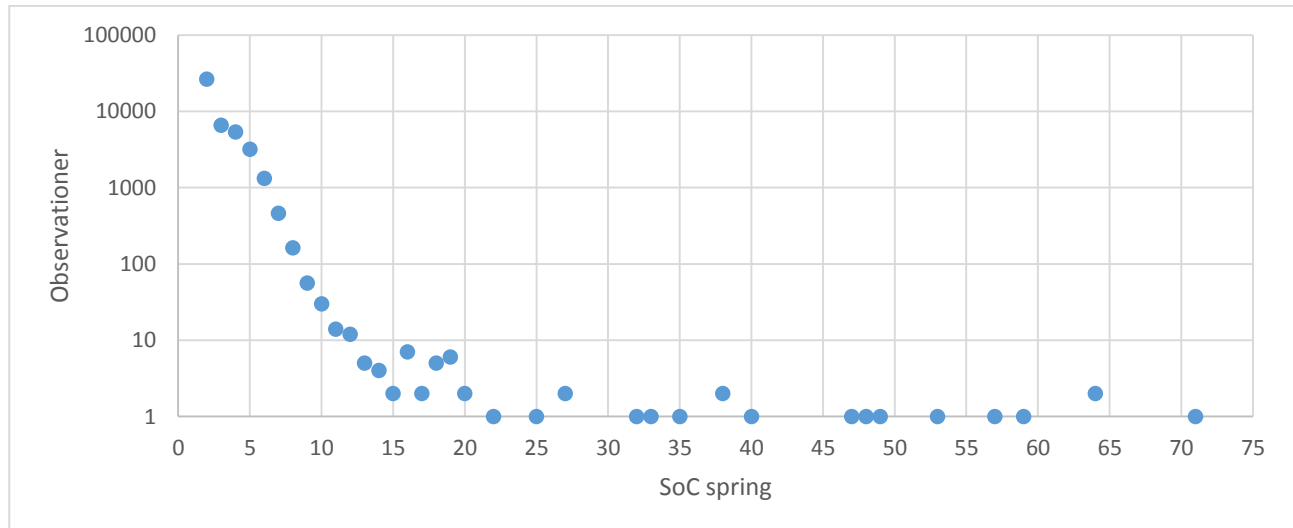
SoC angiver, hvor meget batteriet er opladet i elbilen og under normale omstændigheder rapporteres der en måling for hver gang en opladningsprocent ændrer sig, så opladningsprocenten stiger rimeligt jævnt.

Køretøjs id	UTC tidspunkt	SoC
VF31NZKZZBU801081	2012-10-06 13:42:58,925	78 %
VF31NZKZZBU801081	2012-10-06 13:48:30,523	79 %
VF31NZKZZBU801081	2012-10-06 13:54:01,135	80 %
VF31NZKZZBU801081	2012-10-06 13:59:33,747	81 %
VF31NZKZZBU801081	2012-10-06 14:05:06,350	82 %
VF31NZKZZBU801081	2012-10-06 14:07:42,626	90 %
VF31NZKZZBU801081	2012-10-06 14:08:21,697	91 %
VF31NZKZZBU801081	2012-10-06 14:13:55,305	92 %
VF31NZKZZBU801081	2012-10-06 14:19:25,911	93 %
VF31NZKZZBU801081	2012-10-06 14:24:57,513	94 %
VF31NZKZZBU801081	2012-10-06 14:30:33,117	95 %
VF31NZKZZBU801081	2012-10-06 14:36:06,729	96 %
VF31NZKZZBU801081	2012-10-06 14:41:41,330	97 %
VF31NZKZZBU801081	2012-10-06 14:47:15,932	98 %
VF31NZKZZBU801081	2012-10-06 14:52:51,529	99 %
VF31NZKZZBU801081	2012-10-06 14:58:28,136	100 %

Tabel 7 Eksempel hvor opladningsprocent springer procenter over.

Tabel 7 viser en opladning af køretøjet VF31NZKZZBU801081, hvor opladningsprocenten forøges med 1 for hver måling indtil kl 14:05:06,350. Hver opladningsprocent (SoC) har indtil videre taget 5-6 minutter, men

efter 82 % springer opladningsprocenten til 90 % på under 3 minutter. Den efterfølgende opladningsprocent stiger fra 90 til 91 % på under 1 minut, mens resten af opladningsprocenterne tager 5-6 minutter. Det er uvist, hvorfor dette sker, og hvorvidt dette skyldes batterihåndteringssystemet, en intelligent opladerstander eller andre faktorer.



Figur 5 Antal SoC spring i forhold til hvor mange SoC der springes over.

Figur 5 viser, hvor mange og hvor store SoC spring, der er registreret. Bemærk venligst, at y-aksen er logaritmisk. Det kan ses, at langt de fleste SoC spring ligger på mellem 2 og 8 SoC, hvor der er målt 43.600 spring i dette interval. Tilsvarende er der målt 164 SoC spring på over 8 SoC.

Disse SoC spring er håndteret ved at blive registreret og gemt som en del af opladningen. Derved kan brugeren under analyser bestemme om man vil anvende opladninger, hvor SoC har spring i værdierne.

3.1.2.2 Fluktueringer

Når et køretøj oplades forventes det, at opladningen sker gradvist og at når køretøjet er fuldt opladet forbliver køretøjet i en tilstand, hvor batteriet er 100 % opladet (SoC = 100). Når et batteri er opladet til 100 SoC er opladningen afsluttet og tidspunktet for den fulde opladning kan gemmes. Der er nogle usikkerheder, der gør, at det er svært at fastlægge SoC værdien hen imod slutningen af en opladning.

Køretøjs id	UTC tidspunkt	SoC
VF31NZKZZBU801039	2013-01-05 03:02:10,075	97 %
VF31NZKZZBU801039	2013-01-05 03:06:05,515	98 %
VF31NZKZZBU801039	2013-01-05 03:10:42,033	99 %
VF31NZKZZBU801039	2013-01-05 03:16:07,626	100 %
VF31NZKZZBU801039	2013-01-05 04:09:09,545	99 %
VF31NZKZZBU801039	2013-01-05 04:09:10,547	100 %
VF31NZKZZBU801039	2013-01-05 04:09:34,591	99 %
VF31NZKZZBU801039	2013-01-05 04:09:35,593	100 %
VF31NZKZZBU801039	2013-01-05 04:09:45,610	99 %
VF31NZKZZBU801039	2013-01-05 04:09:46,612	100 %
VF31NZKZZBU801039	2013-01-05 04:09:48,617	99 %
VF31NZKZZBU801039	... 53 gange	99 /100 %
VF31NZKZZBU801039	2013-01-05 04:15:02,193	99 %
VF31NZKZZBU801039	2013-01-05 04:15:10,207	100 %
VF31NZKZZBU801039	2013-01-05 04:15:13,212	99 %
VF31NZKZZBU801039	2013-01-05 04:17:46,489	99 %

Tabel 8 Eksempel på at opladningsprocent fluktuerer mellem 99 og 100 %.

Især den sidste procent kan svinge mange gange mellem 99 og 100 %. Dette kan ses ud fra eksemplet vist i Tabel 8, hvor et køretøj er opladet til 100 % kl. 03:16:07,626, men kl. 04:09:09,545 falder opladningsprocenten til 99 % igen og derefter svinger opladningsprocenten mellem 99 og 100 % indtil kl. 04:17:46,489, hvor opladningen afsluttes. I alt svinger opladningsprocenten i dette tilfælde 63 gange mellem 99 % og 100 % og ofte med få sekunders mellemrum.

Der er ikke umiddelbart nogen forklaring på, hvorfor denne fluktuering sker. Det kan være batterihåndteringssystemet, der har svært ved at vurdere opladningsprocenten og opladningsprocenten derfor ligger og svinger mellem 99 % og 100 %. En anden forklaring kunne være, at noget el udstyr bliver aktiveret og faktisk aflader batteriet, og når batteriet så bliver afladet til 99 % starter opladningen igen, indtil 100 % opladning er nået.

Der er fundet 128.097 tilfælde hvor SoC værdien svinger mellem 99 og 100 SoC. Disse fluktuationer er håndteret ved, at når et batteri når en SoC på 100 % første gang siges opladningen at være fuldendt. Efterfølgende fluktuationer bliver antaget som værende støj og derfor vil tiden, hvor et batteri fluktuerer mellem 99 og 100 % ikke blive en del af en opladning, men blive talt som værende tid, hvor køretøjet er tilsluttet elnettet.

3.2 Rensning af spotpris data

Spotpris data kan hentes fra [3] i et regnearksformat, der er let at anvende og forstå. Det er dog vigtigt at bemærke, at tidspunkterne i regnearket er lokaltid. Dette gør, at når der er overgang til vintertid i Danmark, så er der problemer med manglende data. Dette skyldes, at intervallet 02:00-03:00 findes to gange i lokal tid, men optræder kun én gang i Nord Pool Spots data. Det samme er ikke gældende, når der skiftes fra vintertid til sommertid, her mangler der en time fra 03:00-04:00 i lokaltid.

Denne manglende time fra sommertid til vintertid dækkes ved at tage prisen fra timen før.

3.3 Beregninger på opladningsdata

For at analysere på opladninger er det nødvendigt at beregne flere værdier ud fra opladningsdata, såsom opladningsplaceringer, hvem der oplader og hvornår et køretøj skifter ejer (dette sker cirka hver 3. til 6 måned i projektet Test-en-elbil [2]). Det følgende afsnit vil se på disse definitioner og værktøjer til at bestemme disse værdier.

3.3.1 Brugere og hjemmeladere

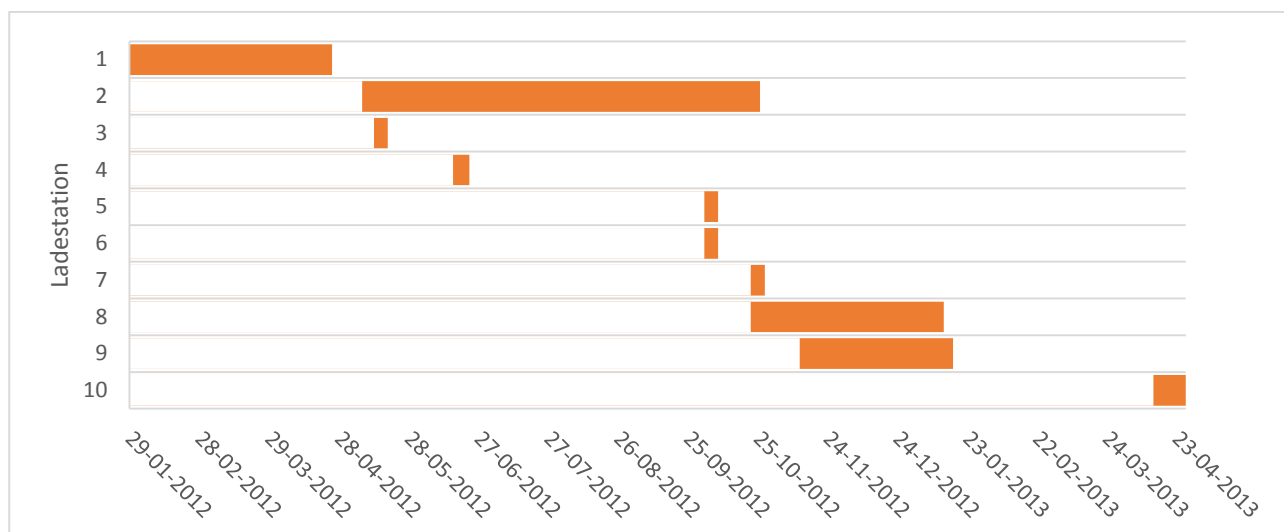
Hver måling er tilknyttet et unikt køretøjs ID, men der er ikke tilknyttet information om hvem der bruger køretøjet. Test-en-elbil projektet går ud på, at brugere kan benytte en elbil i en periode, hvorefter elbilen bliver leveret til en ny bruger. Dette betyder, at hvis man skal kunne skelne mellem brugerne er det nødvendigt at bestemme, hvornår et køretøj skifter til en ny bruger. Dette er ligeledes nødvendigt at bestemme om en opladning foregår hjemme eller ude.

Bestemmelsen af brugerne sker ved at se på de forskellige ladestationer, der er benyttet i løbet af et køretøjs driftsperiode, altså den periode, hvor køretøjet har logget data. Dette sker ud fra en ligelig vægtning mellem, hvor mange gange en standard ladestation er benyttet samt i hvor lang en periode en standard ladestation har været benyttet (antal dage mellem første og sidste opladning ved en ladestation). Denne vægtning kaldes *benyttelsesgraden*.

Antagelsen er, at i perioden hvor en brugeren er tildelt en elbil vil de fleste standard opladninger foregå i hjemmet [2] samt, at brugen af hjemmeopladeren vil strække sig over hele perioden, hvor en bruger har haft tildelt en elbil. Vægtningen mellem antal dage og antal opladninger (*benyttelsesgraden*) giver en opvejning af de to parametre, som hver for sig indikerer et overvejende brug af en ladestation.

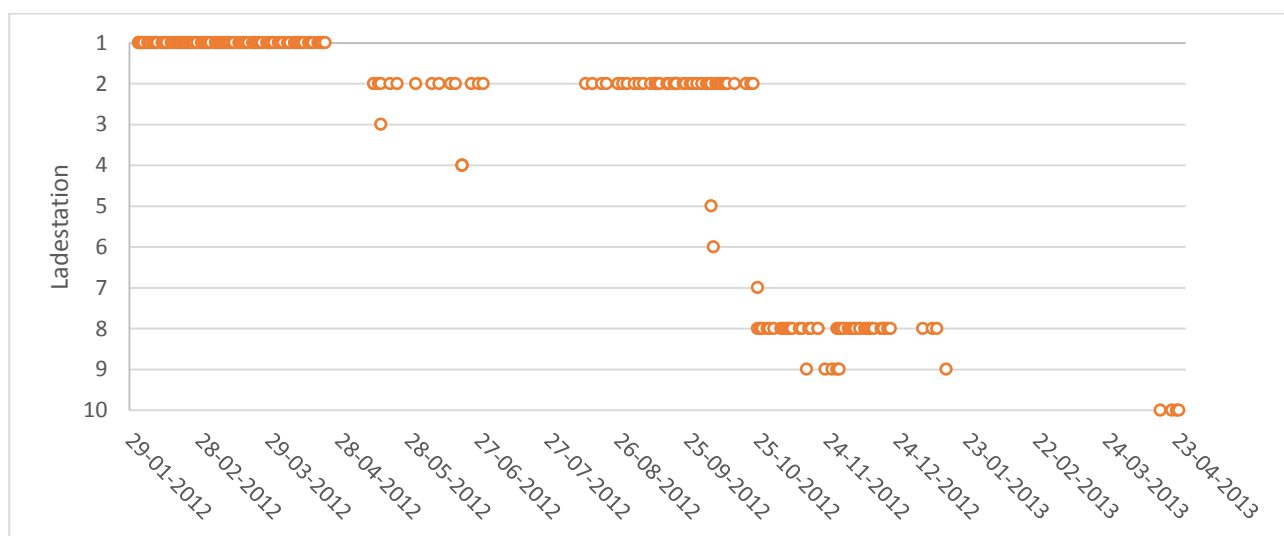
Ladestation	Første dato	Sidste dato	Dage	Opladninger	Benyttelsesgrad
1	2012-02-01	2012-04-22	82	93	175
2	2012-05-11	2012-10-23	166	55	221
3	2012-05-16	2012-05-16	1	1	2
4	2012-06-19	2012-06-20	2	3	5
5	2012-10-05	2012-10-05	1	1	2
6	2012-10-05	2012-10-05	1	1	2
7	2012-10-25	2012-10-25	1	1	2
8	2012-10-25	2013-01-10	78	41	119
9	2012-11-15	2013-01-14	61	6	67
10	2013-04-16	2013-04-24	9	4	11

Tabel 9 Ladestationer benyttet af køretøj med id VF31NZKZZBU801081, sorteret efter første dato.



Figur 6 Ladningsstationer og deres aktive intervaller fra Tabel 9.

Tabel 9 viser ladestationerne benyttet af køretøjet med id VF31NZKZZBU801081. Første kolonne viser et ID for hver ladestation, anden og tredje kolonne viser første og sidste benyttelsesdato for ladestationen, fjerde kolonne viser antal dage mellem første og sidste benyttelsesdato, femte kolonne viser antal opladninger foretaget ved ladestationen og sjette kolonne viser den tidligere nævnte vægtning mellem antal dage og antal opladninger. Figur 6 viser ladestationernes aktive perioder. Her ses det, at nogle ladestationer er benyttet over længere perioder, mens andre kun er benyttet få gange.



Figur 7 Faktiske opladning udført for hver ladestation vist i Figur 6.

Figur 7 viser alle de opladninger, der er foretaget med køretøjet. Sammenlignes med Figur 6 kan det ses, at typisk benyttes en ladestation kun enkelte gange eller også benyttes den løbende. Kun ladestation 2 og 9 har været brugt i skiftende perioder med ugers eller måneders pause.

Det er ukendt, hvor mange brugere, der har haft køretøjet i perioden vist i Figur 6, men ved at se på intervallerne for hvornår ladestationer er benyttet, kan dette udledes.

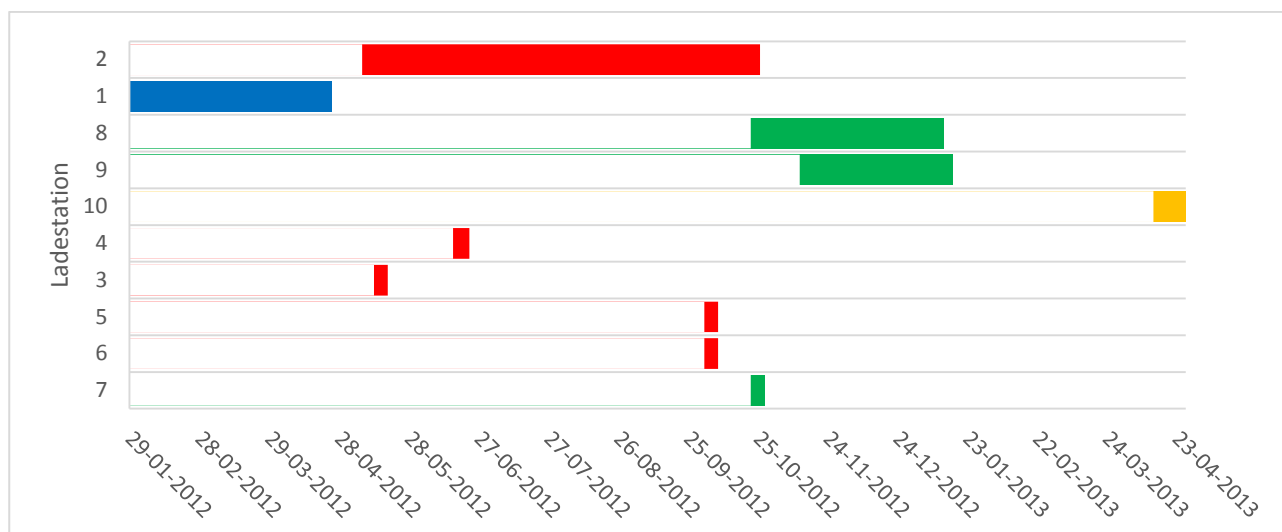
Ladestation	Første dato	Sidste dato	Dage	Opl.	Benyttelsesgrad	Bruger	Tidligere overlap
2	2012-05-11	2012-10-23	166	55	221	A	Nej
1	2012-02-01	2012-04-22	82	93	175	B	Nej
8	2012-10-25	2013-01-10	78	41	119	C	Nej
9	2012-11-15	2013-01-14	61	6	67	C	Delvist
10	2013-04-16	2013-04-24	9	4	11	D	Nej
4	2012-06-19	2012-06-20	2	3	5	A	Ja
3	2012-05-16	2012-05-16	1	1	2	A	Ja
5	2012-10-05	2012-10-05	1	1	2	A	Ja
6	2012-10-05	2012-10-05	1	1	2	A	Ja
7	2012-10-25	2012-10-25	1	1	2	C	Ja

Tabel 10 Ladestationer benyttet af et køretøj (ID=VF31NZKZZBU801081) i 2012, sorteret efter benyttelsesgrad.

Tabel 10 viser den sorterede liste fra Tabel 9, hvor rækkeren er sorteret efter kolonnen *benyttelsesgraden*. Første række viser *ladestation 2*, der har den største benyttelsesgrad på 221. Denne periode bliver herved defineret som værende den periode en bruger har haft køretøjet og perioden bliver tildelt *bruger A*. Den anden række, *ladestation 1*, har en benyttelsesgrad på 175 og da denne periode ikke overlapper med nogle tidligere tildelte perioder til andre brugere, bliver denne periode tildelt *bruger B*. Det samme gør sig gældende for *ladestation 3*, med en benyttelsesgrad på 119, og ikke overlapper med andre brugere, derved bliver perioden tildelt *bruger C*. Disse tre ladestationer bliver bestemt til at være hhv. *bruger A*, *B* og *C*s hjemmeladestationer, da disse ladestationer er de mest benyttede.

Ladestation 9 der anvendes i perioden fra 2012-11-14 til 2013-01-14 overlapper delvist med *ladestation 8*, som strækker sig fra 2012-10-25 til 2013-01-10. *Ladestation 9* er første gang brugt senere end *ladestation 8*, men er også benyttet 4 dage efter sidste opladning ved *ladestation 8* – derfor overlapper de to ladestationer delvist. Men da *ladestation 8* har en højere benyttelsesgrad, primært pga. langt flere opladninger ved *ladestation 8* frem for *9*, hhv. 41 og 6 opladninger og i ringe grade antallet af dage opladninger spænder over, hhv. 78 og 61 dage, bliver opladningerne for *ladestation 9* tildelt *bruger C*, da det vurderes *ladestation 8* er den primære ladestation for *bruger C* og *ladestation 9* er en sekundær ladestation for *bruger C*.

Perioden for *ladestation 10* er ikke tidligere tildelt og derfor bliver denne periode tildelt *bruger D*. Resten af ladestationerne ligger inden for de andre ladestationers perioder og derfor bliver disse tildelt de respektive brugere for perioderne.



Figur 8 Ladestationer sorteret efter Benyttelsesgraden, som vist i Tabel 10. Ladestationerne er farvet efter brugere.

Figur 8 viser en visualisering af ladestationerne for køretøjet, hvor ladestationerne er sorteret efter *benyttelsesgraden*, som i Tabel 10. Ladestationerne er farvede efter brugerne genkendt i Tabel 10, hvor rød er *bruger A*, blå er *bruger B*, grøn er *bruger C* og gul er *bruger D*. Her kan det bl.a. ses, at *ladestation 9* kun delvist overlapper med *ladestation 8*, mens der ikke er noget overlap i ladestationerne for brugerne.

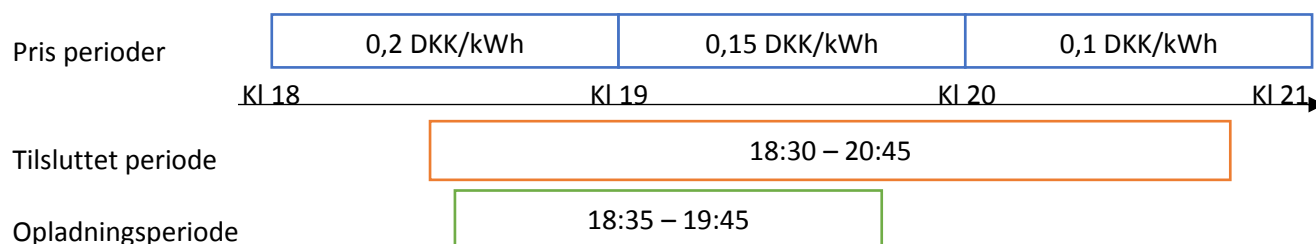
Dernæst bliver alle opladninger med hurtigladere tildelt brugere ud fra hvilke intervaller opladningerne ligger indenfor. Skulle der være hurtigopladninger uden for noget interval mellem aktive brugere (f.eks. hvis en elbil skal flyttes mellem to brugere) findes den nærmeste bruger i forhold til tidspunktet for opladningen og opladningen tilknyttes denne bruger.

3.3.2 Ladestationer

Ladestationer positioner er ikke kendte på forhånd og disse må derfor findes. Når en opladning påbegyndes gemmes alle længdegrader og breddegrader, der rapporteres i løbet af opladningsperioden. Når en opladning er afsluttet beregnes placeringen af opladningen ud fra en trimmet middelværdi af alle længdegraderne og breddegraderne, hvor de mindste og største 10 % er fjernet, for at reducere støj. Dernæst søges der på, hvorvidt der findes en ladestation inden for en radius af 200 meter. Hvis der gør antages det at denne ladestation benyttes, ellers oprettes en ny ladestation.

3.3.3 Opladningsintervaller og priser

For at finde prisen for en opladning benytter vi spotpriserne for hhv. Østdanmark og Vestdanmark. Priserne indlæses som DKK/kWh og priserne er inddelt i timeintervaller. Disse intervaller er bestemt fra data leverandøren og kunne være prisen i DKK/kWh for et vilkårligt interval.



Figur 9 Visualisering af pris-, plugin- og opladningsperioder.

Figur 9 viser et eksempel på en opladning, hvor elpriserne, inddelt i timeintervaller, er vist med de blå kasser. Perioden for, hvornår et køretøj er tilsluttet elnettet er vist med den orange kasse og det kan ses, at køretøjet er tilsluttet over tre prisperioder. Et køretøj kan oplade hele eller dele af tilslutningsperioden, vist med den grønne kasse. Her kan det ses, at fra køretøjet bliver tilsluttet elnettet går der 5 minutter før opladningen begynder og opladningen slutter efter 1:10 time eller 70 minutter. Opladningshastigheden kan ikke forventes at være lineær, derfor er det nødvendigt at se på hver opladningsprocent som en individuel opladning.

Pris perioder	0,2 DKK/kWh	0,15 DKK/kWh	0,1 DKK/kWh	
	KL 18	KL 19	KL 20	KL 21 →
Tilsluttet min.	30	60	45	
Opladningsperioder min.	5 5 10 10	10 15 15		
Opladningsforsinkelse min.	5			

Figur 10 Visualisering af tilpassede opladningsperioder, individuelle opladningsprocenter og ventetid før opladningsstart.

Figur 10 viser, at prisperioderne og den periode et køretøj er tilsluttet elnettet bliver afstemt, så prisperioderne starter og stopper ved tilslutnings start og slut, og tilslutningsperioderne bliver opdelt i henhold til prisperioderne. Ud fra Figur 10 kan det derfor ses, at den første periode starter 18:30 og er 30 minutter lang med prisen 0,2 DKK/kWh. Den anden periode er 60 minutter og prisen er 0,15 DKK/kWh, mens den tredje og sidste periode er 45 minutter lang med en pris på 0,1 DKK/kWh. Opladningsperioden er blevet inddelt i perioder for hver opladningsprocent. Disse opladningsprocenter er ikke afstemt med prisperioderne og en opladningsprocent kan strække sig over flere prisperioder. En ny periode er introduceret her, nemlig en opladningsforsinkelse. Denne forsinkelse bestemmer hvor lang tid der går fra et køretøj er tilsluttet elnettet til en opladning sættes til at starte. Denne variabel er praktisk når den optimale opladningsperiode senere skal findes (se senere).

I det viste eksempel i Figur 10 er den fineste opløsning 1 minut for tilslutningsperioden og opladningsperioderne. Dette er blot for at holde eksemplet simpelt og overskueligt. I praksis arbejdes der med en opløsning på sekund niveau, da en minut granularitet for en opladningsprocent vil være en for grov enhed, især ved hurtigopladere, hvor en opladningsprocent kan tage under et halvt minut.

3.3.3.1 Den faktiske pris

For at finde den faktiske opladningspris tager vi udgangspunkt i eksemplet vist i Figur 10. For at holde udregningen simpel antager vi her, at en opladningsprocent svarer til 0,1 kWh. Dette er *kun* gældende i dette eksempel.

Her vil der først tages udgangspunkt i den første opladningsperiode, som er på 30 minutter. Der er en opladningsforsinkelse før opladningen starter på 5 minutter, derved er der kun 25 minutter tilbage i den opladningsperiode. Den første opladningsprocent tager 5 minutter og opladningsprisen for denne procent er som følgende.

$$1 \text{ opladningsprocent} \times 0,2 \frac{\text{DKK}}{\text{kWh}} \times 0,1 \frac{\text{kWh}}{\text{opladningsprocent}} = 0,02 \text{ DKK}$$

De næste to opladningsprocenter tager hhv. 5 og 10 minutter og passer også i den første opladningsperiode, hvorefter den samlede opladningspris vil være 0,06 DKK og der vil være 5 minutter tilbage i perioden. Den næste opladningsprocent tager 10 minutter, men da der kun er 5 minutter tilbage i den første opladningsperiode bliver denne delt så den del som passer i perioden ($\frac{5}{10}$) bliver betalt i denne periode og resten bliver gemt til næste periode. Derved bliver prisen for denne opladningsprocent for denne periode:

$$\frac{5}{10} \text{ opladningsprocent} \times 0,2 \frac{\text{DKK}}{\text{kWh}} \times 0,1 \frac{\text{kWh}}{\text{opladningsprocent}} = 0,01 \text{ DKK}$$

Da der nu ikke er flere minutter tilbage i den første opladningsperiode startes der på den næste, som indeholder 60 minutter og hvor prisen nu er 0,15 DKK/kWh. Her passer både de resterende 5 minutter fra den tidligere opladningsprocent samt de sidste 10, 15 og 15 minutter og priserne for hver opladningsprocent bliver beregnet som tidligere vist.

3.3.3.2 Den optimale pris

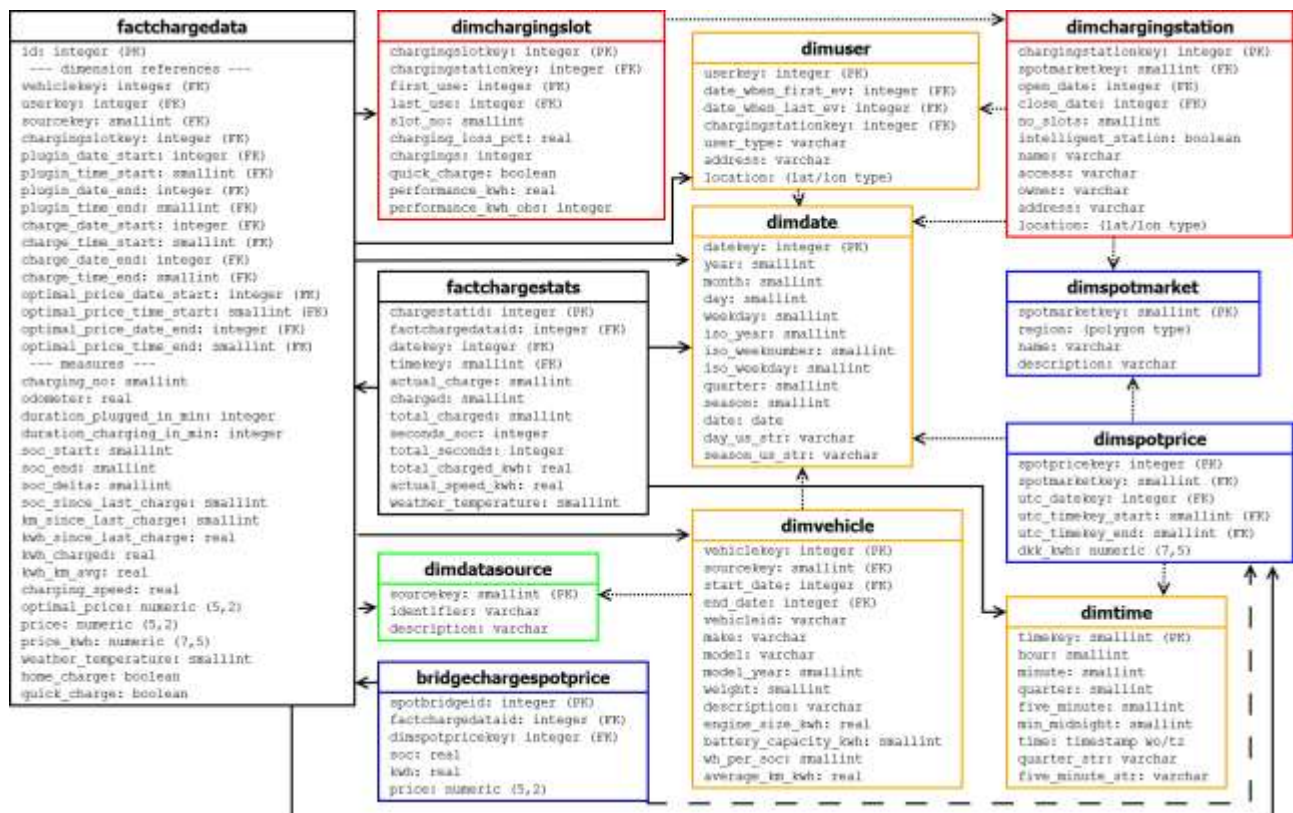
For at finde den optimale opladningspris benyttes samme teknik som vist i afsnit 3.3.3.1. Her ændres opladningsforsinkelsen til at starte med 0 og opladningsprisen findes for de 70 opladningsminutter når de starter samtidigt med tilslutningsperioden. Dernæst øges opladningsforsinkelsen med 1 og dette fortsætter så længe opladningsperioden er fuldstændigt dækket af perioden køretøjet er tilsluttet elnettet. I eksemplet i Figur 10 vil den maksimale gyldige opladningsforsinkelse være 65 minutter (135 minutter (tilslutningsperiode) – 70 minutter (opladningsperiode)).

4 Logisk data-warehouse model

Dette afsnit præsenterer den logiske data model for et data warehouse [4] til at gemme opladningsdata og elpris data. Modellen er fuld implementeret og indeholder alle opladningsdata fra Test-en-elbil projekt tilgængeligt samt alle elpriser fra Nord Pool Spot for årene 2012 og 2013. Rapporten vil i det efterfølgende afsnit være meget teknisk betonet. Der er indsat referencer til bogen "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" [4], som er en anerkendt og forholdsvis let tilgængelig bog om data warehouse modeller.

4.1 Model

Den logiske model er lavet som et stjerneschema [4] med visse snowflakede dimensioner [4]. Figur 11 viser designet af den logiske model, hvor der er to fact tabeller [4] (sorte), en bridge tabel [4] mellem en fact tabel og en dimension [4] (mørke blå) og ni dimensioner (farvet efter deres egenskaber). De hele linjer viser referencer [4] fra fact tabeller, de stiplede linjer referencer fra bridge tabel og de prikkede linjer referencer mellem dimensioner.



Figur 11 Logisk model for opladningsdata warehouse.

4.2 Dimensioner

Der er i alt 9 dimensioner, som er vist med forskellige farver i Figur 11. De gule dimensioner beskriver informationer om dato, tidspunkt, køretøj og forskellige brugere af et køretøj. Den grønne dimension beskriver metadata om datakilden. De blå dimensioner beskriver informationer om elmarkedet og elpriser, bl.a. spatiale informationer om de geografiske grænser for elmarkedet. De røde dimensioner beskriver informationer om ladestationer og ladepladser, bl.a., den geografiske placering.

4.2.1 Dimdate

Dimdate dimensionen indeholder standard informationer om datoer samt informationer såsom ISO år, ugenumre og ugedage, hvor den sidste uge af et år er en hel uge og et år altid starter en mandag. I alt er der 12 attributter, og dimensionen er ikke tidszone afhængig, altså kan den refereres både som værende et UTC tidspunkt og et lokalt tidspunkt. Der er flere hierarkier i dimensionen [4], f.eks. ISO hierarkiet, årstider og kvartaler. Den primære nøgle er en såkaldt smartnøgle [4], sammensat af år, måned og dag, så datoen 2014-11-21 vil have nøglen 20141121.

4.2.2 Dimtime

Dimtime dimensionen indeholder alle tidspunkter i et døgn. Granulariteten er ned til 1 minut og dimensionen er ikke knyttet til nogen tidszone. Der er i alt 8 attributter, som gemmer time og minut, samt opdelinger af døgnet i kvarterer og fem-minutters intervaller. Herudover er der kontekstuelle repræsentationer af disse tidsopdelinger. Den primære nøgle er en smartnøgle, sammensat af time og minuttal, så tidspunktet 19:43 vil have smartnøglen 1943.

4.2.3 Dimvehicle

Dimvehicle dimensionen indeholder informationer om hvert unikt køretøj og har 13 attributter. Et køretøj er identificeret ud fra en *sourcekey* og et *vehicleid*. Der er to referencer til *dimdate*, da det er registreret, hvornår køretøjet første og sidste gang er set i brug. Derudover er der lagret generelle informationer om køretøjet og dets egenskaber. Den primære nøgle er en såkaldt surrogat nøgle [4], der giver en unik nøgle og intet beskriver om køretøjet.

4.2.4 Dimuser

Dimuser dimensionen beskriver en individuel bruger af et køretøj og har 8 attributter. En attribut registrerer et koordinat og en anden den tilsvarende postadresse som værende brugeres hjem og brugeren er tilknyttet én *dimchargingstation*. Denne ladestation er brugeres hjemmelader. En bruger er beskrevet med en type, som kan være privat, hjemmehjælp, taxi eller andet. Der registreres hvilken dato brugeren første og sidste gang er bestemt til at have brugt en elbil.

4.2.5 Dimdatasource

Dimdatasource dimensionen beskriver, hvorfra data er modtaget, da flere datakilder er tilgængelige for systemet. Datakildedimensionen har to attributter et ID og en beskrivelse.

4.2.6 Dimspotmarket

Dimspotmarket dimensionen beskriver et spotmarked [3], som er en spatial region, hvor elpriserne er de samme indenfor. Spotmarked dimensionen har tre attributter, nemlig en spatial polygon, der definerer området, et navn samt en beskrivelse.

4.2.7 Dimspotprice

Dimspotprice dimensionen beskriver, hver prisperiode for hvert spotmarked. En spotprisperiode er defineret ud fra et spotmarked, en dato samt et start og et slut tidspunkt.

4.2.8 Dimchargingstation

Dimchargingstation beskriver en enkelt ladestation. En ladestation er tilknyttet et spotmarked og der er 10 attributter til en *dimchargingstation*. Disse attributter beskriver navn, ejer, postadresse, koordinater, antal opladningsslots, åbnings og lukningsdato og informationer om hvorvidt ladestation er intelligent.

4.2.9 Dimchargingslot

Dimchargingslot beskriver den enkelte ladeplads i *dimchargingstation*. Et opladningsslot har 9 attributter, som beskriver, hvornår opladningsslottet er benyttet, hvor mange gange, med hvilken hastighed der oplades og lignende informationer.

4.3 Fact tabeller

Der er i alt to fact tabeller i data warehouse. Disse to tabeller gemmer informationer om opladningerne og disse refererer dimensionerne, som understøtter fact tabellerne.

4.3.1 Factchargedata

Factchargedata indeholder en række for hver individuel opladning, der er foretaget. Der er 16 referencer til dimensioner og 19 measures [4], der beskriver hver enkelt opladning. 12 af referencerne beskriver, hvornår et køretøj har været tilsluttet elnettet, hvornår køretøjet har opladet samt hvornår køretøjet optimal kunne have opladet. En tæller gemmer, hvor mange gange et køretøj har været opladet og der gemmes informationer om, hvor mange minutter et køretøj har været tilsluttet elnettet og hvor mange minutter et køretøj har ladet op.

Den faktiske opladningspris er gemt samt den optimale opladningspris dvs. hvis der var blevet opladet i et optimalt tidsrum. Der er også gemt informationer om, hvorvidt opladningen forgår hjemme eller ved en hurtigoplader og informationer om udendørstemperaturen på tidspunktet, hvor opladningen er startet.

4.3.2 Factchargestats

Factchargestats gemmer en serie målinger for hver opladning. For hver gang der er registreret én opladning i *factchargedata* er der i *factchargestat* gemt en række for hvert opladningsspring, der er ladet op i løbet af en opladning. Et opladningsspring er som regel en enkelt procentændring (en SoC), men ved spring i opladningsprocenter gemmes en række for hele springet. *Factchargestats* refererer *factchargedata* og for hver opladningsprocent er der gemt informationer om det antal sekunder det tog at oplade en procent (en SoC), opladningshastigheden, den akkumulerede tid for hele opladningen og informationer om udendørstemperaturen.

4.4 Bridge tabel

Der er en bridge tabel, som gør det muligt at have en 1-til-mange relation mellem tabeller. Denne bridge tabel sørger for, at én opladning i *factchargedata* kan referere alle de opladningsperioder, der er benyttet fra *dimchargingslot* dimensionen.

4.4.1 Bridgechargespotprice

Bridgechargespotprice tabellen refererer alle de prisintervaller der er benyttet fra *dimspotprice* i løbet af en opladning. For hver opladning i *factchargedata* er der, for hvert benyttet opladningsinterval fra *dimspotprice* en række, der beskriver, hvor meget der er opladet i det givne interval samt prisen betalt i dette opladningsinterval.

4.5 Implementation

Hele data warehouse designet beskrevet ovenfor, og vist i Figur 11, er implementeret vha. PostgreSQL version 9.3 [11]. De yderligere renseregler, der er beskrevet i denne rapport, er implementeret i programmeringssproget Python version 3.4 [12].

Tabel	Rækker	Tabel størrelse	Index størrelse	Total størrelse
Dimspotmarket	2	16 KB	24 KB	40 KB
Dimdatasource	17	16 KB	32 KB	48 KB
Dimvehicle	176	32 KB	64 KB	96 KB
Dimdate	730	56 KB	368 KB	424 KB
Dimuser	807	104 KB	40 KB	144 KB
Dimtime	1.440	98 KB	416 KB	514 KB
Dimchargingstation	1.905	216 KB	64 KB	280 KB
Dimchargingslot	1.905	144 KB	128 KB	272 KB
Factchargedata	57.511	9 MB	6 MB	15 MB
Dimspotprice	142.698	12 MB	26 MB	38 MB
Bridgechargespotprice	256.280	15 MB	22 MB	37 MB
Factchargestats	2.474.013	181 MB	106 MB	287 MB
Total	2.937.484	218 MB	161 MB	379 MB

Tabel 11 Størrelse af tabeller i antal rækker og bytes.

Fact tabeller, bridge tabeller og dimensioner er alle implementeret som tabeller i PostgreSQL og størrelsen af de enkelt tabeller kan ses i Tabel 11. Her er det tydeligt, at *factchargestats* er klart den største tabel. Dette er ikke nogen overraskelse, da denne gemmer en måling for hver opladespring for hver opladning.

	Fact tabel	Factchargedata	Dimensioner	Dimdate	Dimtime	Dimvehicle	Dimuser	Dimdatasource	Dimspotmarket	Dimspotprice	Dimchargingstation	Dimchargingslot
Fact tabeller												
Factchargedata				X	X	X	X	X				X
Factchargestats		X		X	X							
Bridge tabel												
Bridgechargespotprice		X								X		
Dimensioner												
Dimvehicle				X				X				
Dimuser				X							X	
Dimspotprice				X	X				X			
Dimchargingstation				X					X			
Dimchargingslot				X							X	

Tabel 12 Enterprise Bus Matrix af data warehouse modellen fra Figur 11.

Tabel 12 viser en Enterprise Bus Matrix [4] for den logiske model fra Figur 11. Her ses relationerne mellem fact tabeller, bridge tabeller og dimensioner

5 Konklusion

Denne rapport har beskrevet, hvordan information om opladning af elbiler er fundet ud fra GPS og CAN bus målinger fra projektet Test-en-elbil. Denne information om opladningerne er integreret med information om de historiske elpriser og gemt i et data warehouse.

GPS og CAN bus målingerne er fra 2012 og 2013 og opsamlet fra 176 meget ens biltyper Citroen C-Zero, Mitsubishi iMiev og Peugeot Ion. Typiske fejl med uventede værdier for opladning er dokumenteret og data er rensset for disse fejl. Data om elpriser er hentet fra Nord Pool Spot. Disse data er rensset og tilrettet for mindre fejl. Der er udelukkende data fra projektet Test-en-elbil så det kan ikke vurderes om der er mange eller få fejl i opladningsdata.

Selve integrationen af data om opladningen og data om elpriserne er foretaget i et data warehouse. Dette warehouse er fuldt implementeret på en PostgreSQL platform. Al data for 2012 og 2013 om opladning af elbiler og elpriser er indlæst i data warehouset.

Designet af data warehouset er generisk, hvilket betyder, at det er forholdsvis simpelt at indlæse data fra andre kilder og udvide data grundlaget til også at dække andre lande.

Litteraturliste

- [1] Clever, »Sådan læses filerne«.
- [2] Clever, »Test-en-elbil afsluttende rapport,« 2014.
- [3] Nord Pool, [Online]. Available: www.nordpoolspot.com.
- [4] R. Kimball og M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, John Wiley & Sons, 2013.
- [5] O. Andersen, B. B. Krogh og K. Torp, »Analyse af elbilers forbrug for perioden 2012-2013,« DBTR-36 Aalborg University, 2014.
- [6] O. Andersen, B. B. Krogh og K. Torp, »Analyse af elbilers forbrug,« DBTR-34 Aalborg Universitet, 2013.
- [7] B. B. Krogh, O. Andersen og K. Torp, »Electric and Conventional Vehicle Driving Patterns,« i *ACM SIGSPATIAL*, 2014.
- [8] B. B. Krogh, O. Andersen og K. Torp, »Analyzing Electric Vehicle Energy Consumption using Very Large Data Sets,« i *Submitted for publication*, 2014.
- [9] O. Andersen, B. B. Krogh, C. Thomsen og K. Torp, »An Advanced Data Warehouse for Integrating Large Sets of GPS Data,« *DOLAP*, pp. 13-22, 2014.
- [10] Ing.dk, »Overskudsstrøm gav negative elpriser i julen,« 04 01 2013. [Online]. Available: <http://ing.dk/artikel/overskudsstrom-gav-negative-elpriser-i-julen-135324>.
- [11] PostgreSQL, »PostgreSQL Homepage,« [Online]. Available: postgresql.org.
- [12] Python, »Python Homepage,« [Online]. Available: python.org.

Bilag 1

Der er i GPS og CAN bus målingerne fundet et antal steder for kilometertallet i data hopper fra 32.768 til 4.294.934.528 km. Det antages, at sammenhængen mellem disse tal kan forklares på følgende måde. Det næste tal for kilometertælleren kan udregnes ud fra formelen

$$\text{høj kilometertællerværdi} - 2^{32} + 2^{16} = \text{kilometertæller}$$

Og hvis man indsætter kilometertallet, vil man få

$$4.294.934.528 - 2^{32} + 2^{16} = 32.768$$

Altså ser det ud til, at det er muligt at udregne den faktiske kilometertæller, men hvorvidt denne udregning er pålidelig er uvist. Der er ikke registreret nogle kilometertæller målinger mellem 32.768 og 4 milliarder, så dette antages at være et generelt problem for alle køretøjer.

I alt er der registreret 2.716.024 målinger, hvor kilometertælleren er over 4 milliarder.

Antager vi, at kilometertælleren gemmes som en *16 bit signed integer*, dvs. et positivt/negativt heltal der kan repræsenteres ved 16 bits, da vil en variabel kunne gemme værdier mellem -32.768 og 32.767. Hvis man laver regnestykket $32.767 + 1$ da vil dette, når vi regner med *16 bit signed integers*, resultere i -32.768. Binært vil regnestykket se således ud: "01111111 11111111" + "1" = "10000000 00000000". Dette resultat ligner stadig ikke vores kilometertæller på over 4 milliarder. Men hvis kilometertælleren bliver omdannet til en *32 bit unsigned integer*, dvs. et positivt heltal der kan repræsenteres ved 32 bits, og vil tilføjer en foranstillet padding med 1 bits til de 16 ekstra bits, da vil det nye tal komme til at se således ud: "11111111 11111111 **10000000 00000000**" hvor de fede bits er tallet der konverteres og de kursive bits er padding, da vil vores decimale repræsentation, ved antagelsen at tallet nu er et *32 bit unsigned integer* være 4.294.934.528.

Hvorvidt dette beskriver, hvad der faktisk sker vides ikke, men resultatet passer med kilometertælleren.